Report on Sensor Data Analysis and Model Performance

Overview

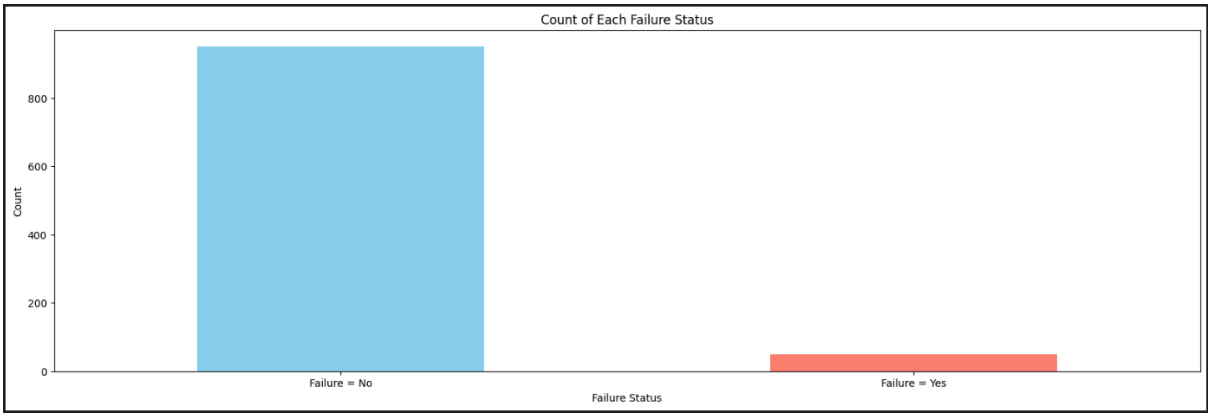
This report presents a comprehensive analysis of a sensor dataset containing 1,000 records. The dataset includes various sensor readings over time, with an emphasis on evaluating the relationships between sensor measurements and failure occurrences. The analysis encompasses the distribution characteristics of sensor variables, correlation with failure rates, and the performance of a Random Forest model trained on this data.

Dataset Summary

The dataset provides sensor data across multiple variables:

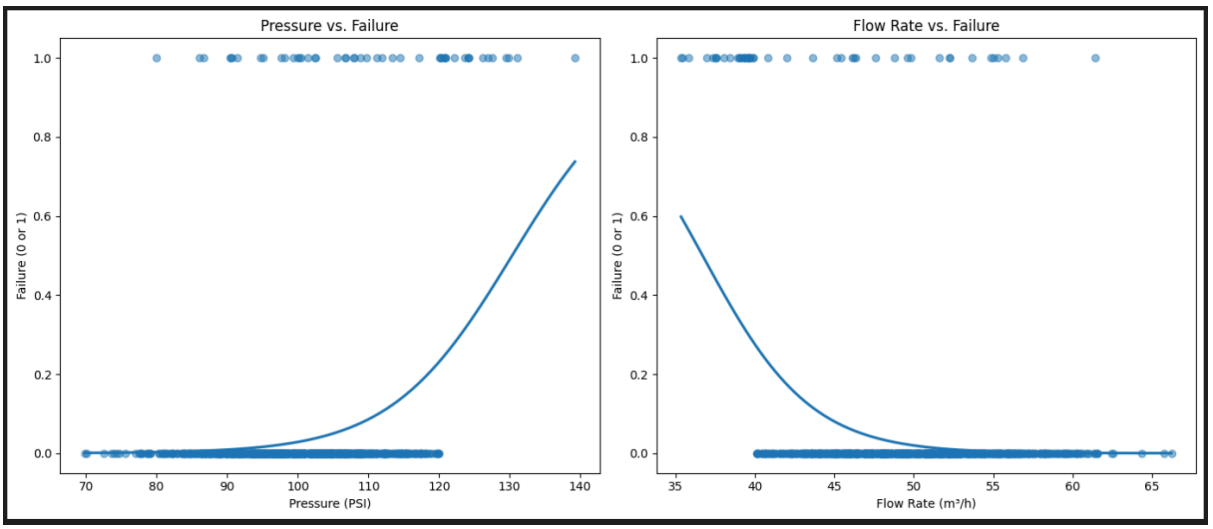
- **Vibration Level**: Average of 0.502
- **Temperature**: Approximately 70.35°C
- **Pressure**: Around 100.06 PSI
- **Flow Rate**: 49.91 m³/h

The failure rate in the dataset is notably low, occurring in only 4.9% of the records.



Correlation Analysis

- **Pressure (PSI) and failure** exhibit a moderate positive correlation (0.236). This suggests that an increase in pressure is associated with a slight increase in the likelihood of failure.
- **Flow Rate (m³/h) and failure** show a moderate negative correlation (-0.275). This indicates that higher flow rates are associated with a reduced likelihood of failure.

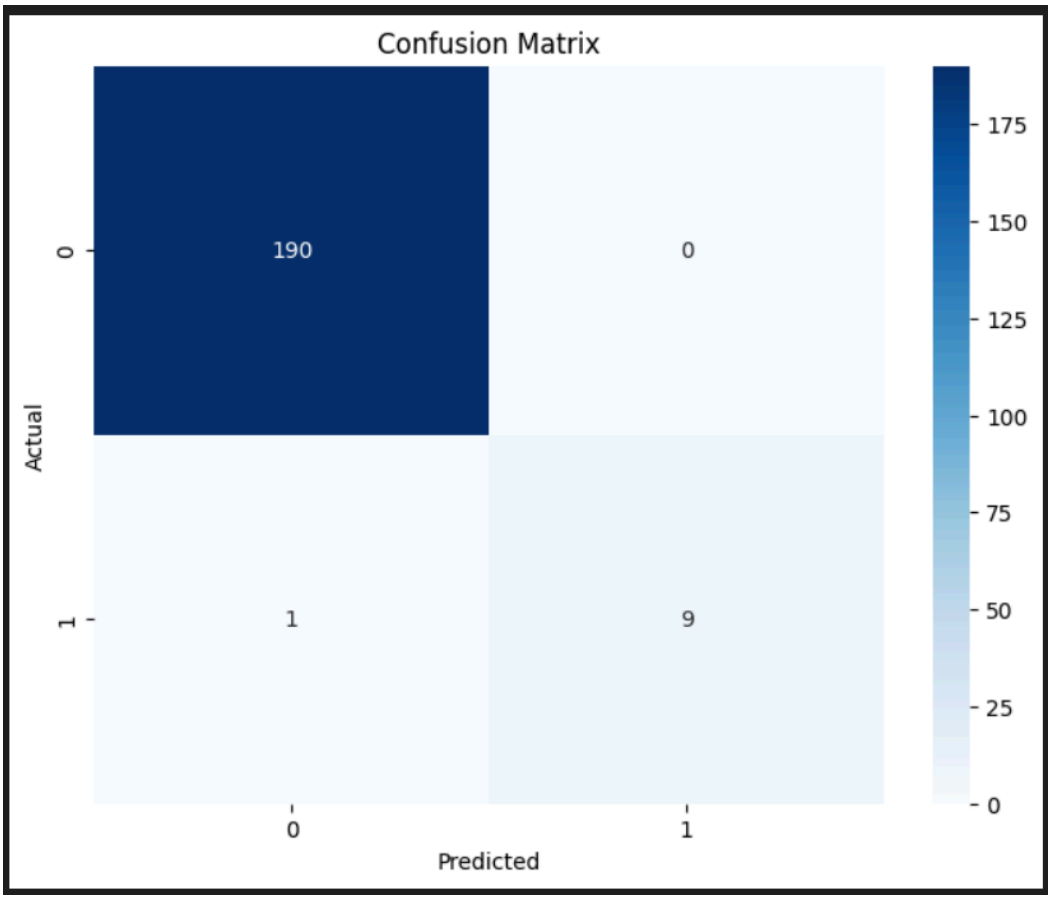


Distribution Characteristics

1. **Vibration Level**
    - **Kurtosis**: Close to 0, indicating a distribution that is relatively normal.
    - **Skewness**: Also close to 0, suggesting a fairly symmetric distribution.
2. **Temperature (°C)**
    - **Kurtosis**: Near 0, which signifies a distribution that is neither excessively peaked nor flat, aligning with a normal distribution.
    - **Skewness**: Very close to 0, showing a symmetric distribution.
3. **Pressure (PSI)**
    - **Kurtosis**: Slightly positive but close to 0, indicating tails similar to those of a normal distribution.
    - **Skewness**: Minimal, suggesting a symmetric distribution.
4. **Flow Rate (m³/h)**
    - **Kurtosis**: Slightly negative, implying a distribution with marginally lighter tails compared to a normal distribution, though still close to 0.
    - **Skewness**: Almost 0, indicating symmetry.

Model Performance

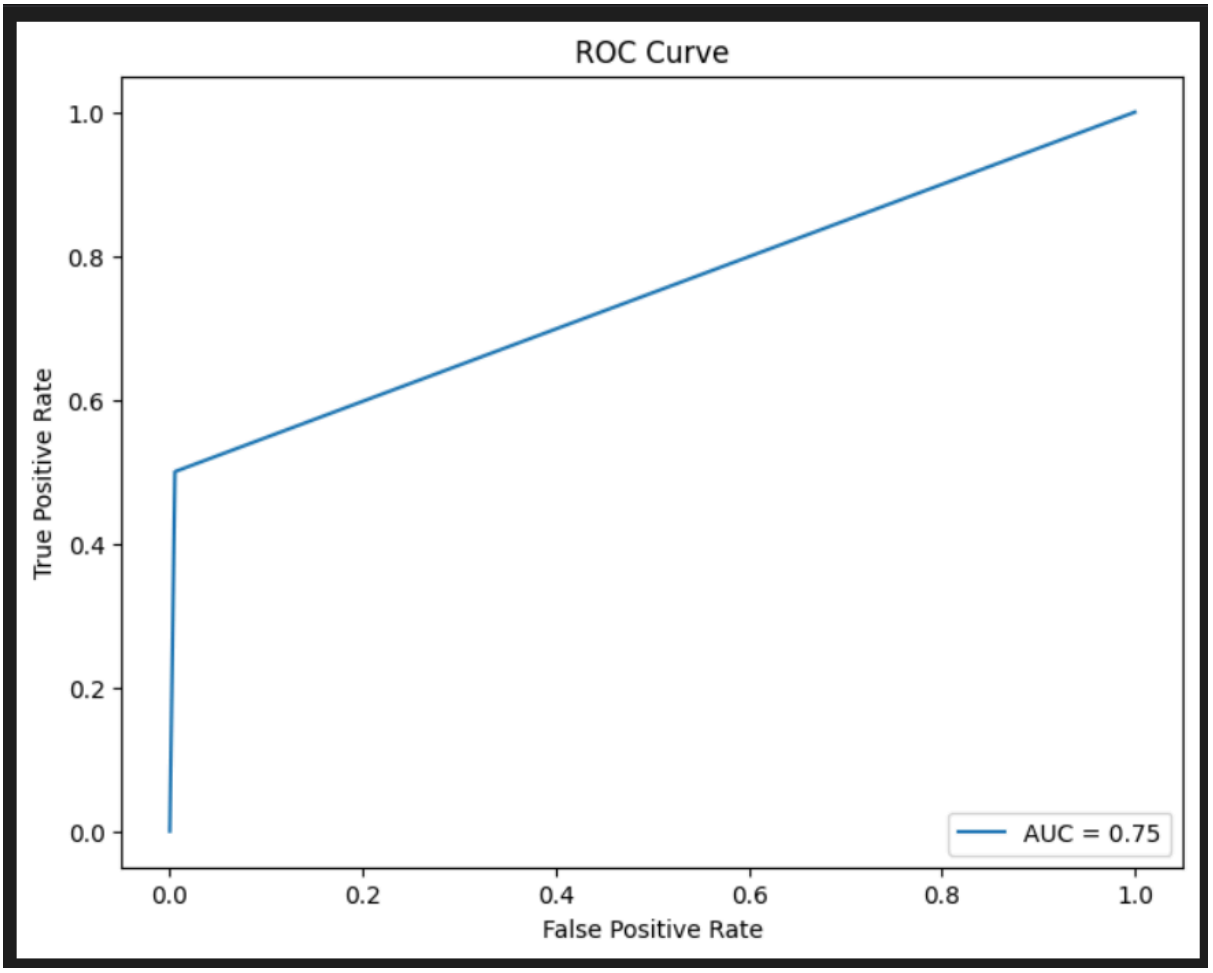
A Random Forest model has been trained on the dataset, achieving an accuracy of 95%. The model's performance is further evaluated using the confusion matrix, which is illustrated in the graph below.



Model Performance on Forecasting: LSTM

The performance of the Long Short-Term Memory (LSTM) forecasting model was assessed, and the results indicate that the model struggled with class imbalance in the dataset. Below is the detailed classification report for the LSTM model:

	Precision	Recall	F1-Score	Support
Class 0	0.97	0.99	0.98	190
Class 1	0.83	0.50	0.62	10
Accuracy			0.97	200
Macro Avg	0.90	0.75	0.80	200
Weighted Avg	0.97	0.97	0.97	200



Conclusion

The dataset analysis reveals well-behaved distributions for the sensor variables and provides insights into their relationships with failure occurrences. The Random Forest model demonstrates a high accuracy rate, indicating its effectiveness in predicting failures based on the sensor data. The detailed analysis and model performance underscore the robustness of the dataset and the effectiveness of the predictive model.